

## University of Groningen

### Random effects models for directed graphs with covariates

Zijlstra, Bonne Jelle Hepke

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2008

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Zijlstra, B. J. H. (2008). *Random effects models for directed graphs with covariates*. [Thesis fully internal (DIV), Rijksuniversiteit Groningen]. s.n.

#### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

---

## Samenvatting (Summary in Dutch)

Het  $p_2$  model is een statistisch model voor de analyse van dichotome sociale netwerkdata met behulp van verklarende variabelen. Een sociaal netwerk wordt gevormd door een groep actoren en hun onderlinge relaties. In de dichotome sociale netwerk data waar het  $p_2$  model voor bedoeld is, wordt voor elk paar van actoren slechts aangegeven of de gerichte relaties tussen hen (van de ene actor in het paar naar de andere actor en vice versa) aan- of afwezig zijn. Afhankelijk van de aanwezige relaties zijn er voor elk paar van actoren (elke dyade) dus vier uitkomsten mogelijk: een lege dyade zonder relaties, twee asymmetrische dyades met een enkele relatie en een wederkerige dyade met twee relaties. Het  $p_2$  model modelleert de kansen voor deze vier mogelijke dyadische uitkomsten. Het model bevat parameters voor de algehele dichtheid van relaties, de reciprociteit van de relaties binnen de dyades, en random zender en ontvanger effecten. Met deze random effecten wordt heterogeniteit tussen actoren in het zenden en ontvangen van relaties en de eventuele afhankelijkheid hiertussen gemodelleerd. Voor alle parameters kunnen verklarende variabelen in het model worden opgenomen. Het  $p_2$  model kan dus beschouwd worden als een bijzonder model voor multinomiale regressie, of als een bivariaat logistisch regressiemodel, met gekruiste random effecten.

De combinatie van de dichotome observaties met de gekruist geneste random effecten stelt extra eisen aan de schattingsprocedure voor het  $p_2$  model. Om deze reden worden in hoofdstuk 2 drie Markov Chain Monte Carlo (MCMC) algoritmes geïntroduceerd om de parameters van het  $p_2$  model te schatten, ter vervanging van een eerder ontwikkeld Iterative Generalized Least Squares (IGLS) algoritme. Twee van de MCMC algoritmes selecteren voorstellen voor parameterwaarden die gegenereerd zijn door middel van een zogenaamde 'random walk'. Het derde algoritme is een 'independence chain sampler' waarin onafhankelijke voorstellen voor parameterwaarden worden gebruikt. In dit laatste algoritme

en in één van de random walk algoritmes wordt bij het genereren van voorstellen voor parameterwaarden gebruik gemaakt van een normale benadering voor de gezamenlijk verdeling van waarnemingen, random effecten en parameters, verkregen met behulp van de verwachting en covariantiematrix onder het  $p_2$  model.

In een grootschalig simulatieonderzoek worden de MCMC schattingen met de IGLS schattingen vergeleken. De resultaten laten zien dat de MCMC schattingen slechts een kleine systematische afwijking vertonen en dat de geschatte intervallen de gesimuleerde parameterwaarden zo vaak bevatten als verwacht, terwijl dit voor de IGLS schattingen niet het geval is. Hoewel de resultaten voor alle MCMC algoritmes vergelijkbaar zijn, is er om praktische redenen voor gekozen om één specifiek algoritme te gebruiken voor een multilevel en een multivariate uitbreiding van het  $p_2$  model. Dit is het random walk algoritme dat de normale benadering van de dichotome netwerkdata toepast.

In hoofdstuk 3 wordt een multilevel  $p_2$  model voorgesteld voor meerdere onafhankelijke netwerken die verondersteld worden representatief te zijn voor een bepaalde populatie. Hierin wordt een identiek  $p_2$  model gespecificeerd voor alle geobserveerde netwerken, met parameters die kunnen variëren over de verschillende netwerken. Het model wordt geïllustreerd aan de hand van data die betrekking hebben op ontvangen praktische steun tussen Nederlandse middelbare schoolleerlingen van verschillende ethnische achtergronden.

Een multivariaat  $p_2$  model voor de analyse van verschillende gerichte dichotome netwerken, geobserveerd bij dezelfde groep van actoren, wordt voorgesteld in hoofdstuk 4. Het model omvat parameters voor de bivariate samenhang tussen meerdere afhankelijke netwerken. Deze parameters representeren multiplexiteit, uitwisseling, gedeeltelijke bivariate reciprociteit en volledige bivariate reciprociteit. Voor deze extra parameters kunnen verklarende variabelen worden toegevoegd, net als voor de oorspronkelijke  $p_2$  parameters die ook in het model aanwezig zijn. In het multivariate model wordt de afhankelijkheid tussen de geobserveerde relaties op actor niveau ook met random effecten gemodelleerd, maar nu inclusief de associaties tussen de afhankelijke netwerken. Door het gebruik van random effecten en de mogelijkheid

voor het opnemen van covariaten ontstijgt dit model bestaande log-lineaire modellen voor meerdere afhankelijke netwerken. Het multivariate  $p_2$  model wordt geïllustreerd met drie toepassingen: één waarbij twee herhaalde metingen van vriendschapsnetwerken van eerstejaars studenten worden geanalyseerd, één waarbij naar de betrouwbaarheid van gerapporteerde praktische steun tussen adolescenten op een middelbare school wordt gekeken, en één waarbij de samenhang tussen status enerzijds en advies, samenwerken en vriendschap anderzijds wordt onderzocht met data van een Amerikaans advocatenkantoor.

Hoofdstuk 5 behandelt Bayesiaanse methoden voor modelselectie voor het  $p_2$  model. Verschillende benaderingen van Bayes factoren worden empirisch onderzocht. Voor een netwerk in een Nederlandse middelbare school wordt onderzocht of leerlingen meer emotionele steun ervaren vanuit de eigen ethnische groep. De benaderde Bayes factoren lijken goed bruikbaar, maar er moet wel rekening worden gehouden met met een redelijk grote foutenmarge.

Alle hiervoor genoemde ontwikkelingen (de MCMC schattingsmethoden, het multilevel en het multivariate  $p_2$  model en verschillende indices voor Bayesiaanse modelselectie) zijn geïmplementeerd in de  $p_2$  module in StOCNET (Boer et. al, 2006). Dit is een gratis verkrijgbaar softwarepakket voor de statistische analyse van sociale netwerk data.

De beschikbaarheid van de nieuwe  $p_2$  modellen en software heeft al verschillende onderzoeken gefaciliteerd. Enkele van de publicaties waarin deze zijn gebruikt, zijn de volgende. Keating, Ayanian, Cleary en Marsden (2007) hebben met het  $p_2$  model onderzoek gedaan naar predictoren voor invloedrijke (advies)gesprekken onder artsen. Veenstra et al. (2007) hebben het multilevel  $p_2$  model toegepast op 54 netwerken van (zelf gerapporteerd) pesten en gepest worden, waarbij het effect van verklarende variabelen op actor en dyadisch niveau zijn onderzocht. Tot slot hebben Baerveldt, Zijlstra, De Wolf, Van Rossem en Van Duijn (2007) het multilevel  $p_2$  model toegepast om ethnische grenzen in vriendschapsnetwerken op Nederlandse en Vlaamse middelbare scholen met elkaar te vergelijken.